Appendix 29. Model Archival Summary for Enterococci Concentration at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary summarizes the Enterococci concentration (Ent) model developed to compute 15-minute Ent from July 19, 2012 onward. This model supersedes all previous models.

Site and Model Information

Site number: 06892350
Site name: Kansas River at De Soto, Kansas
Location: Lat 38°59'00", long 94°57'52" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 SE 1/4 sec.28, T.12 S., R.22 E., Leavenworth County, KS, Hydrologic Unit 10270104.

Equipment: An YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll was installed from August 2012 through May 2014. From June 2014 to the present (2015) a Xylem YSI EXO2 water-quality monitor equipped was installed with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll. The monitor is housed in a 4-inch diameter galvanized steel pipe. Readings from the water-quality monitor are recorded every 15 minutes and transmits data by way of satellite, hourly.

Date model was created: October 15, 2015

Model calibration data period: July 19, 2012 – June 29, 2015

Model application date: July 19, 2012 onward

Model-Calibration Dataset

All data were collected using U.S. Geological Survey (USGS) protocols and are stored in the National Water Information System (NWIS) database. Linear regression models were developed using the open-source software package "R." Explanatory variables selected as inputs to linear regression were physicochemical properties: specific conductance, pH, water temperature, dissolved oxygen, turbidity, chlorophyll fluorescence, and streamflow. Seasonal components (sine and cosine variables) were also evaluated as explanatory variables in the models to determine if seasonal changes affected the model. All combinations of physicochemical properties and a seasonal component were evaluated to determine which combinations produced the best models.

The final selected regression model is based on 50 concurrent measurements of Ent concentration and turbidity (Turb) collected from July 19, 2012 through June 29, 2015. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Twenty-four samples were estimated due to colony counts outside of the ideal range. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of that range are considered potential outliers and are investigated. One sample, June 11, 2014, was found to have potential errors in collection and processing, and has been removed from the dataset. All other potential outliers were not found to have errors associated with collection, processing, or analysis, and were therefore considered valid.

Enterococci Sampling Details

Cross-section samples are typically collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The equal-width-increment (EWI) method is used, and samples typically are composited for analysis. Cross-section samples are collected every 2 weeks from March through October, once a month from November through February, and during selected runoff events. A FISP US DH-95, D-95, or D-96A1 depth integrating sampler is used from the bridge; and a DH-81 or DH-95 hand sampler is used for boat samples. Samples are analyzed for Ent concentration at the USGS Kansas Water Science Center (KSWSC) Lab in Lawrence, Kansas.

Model Development

Regression analysis was done using R by examining Turb, streamflow, and other continuously measured data as explanatory variables for estimating Ent concentration. A variety of models that predict Ent, $(Ent)^2$, $\sqrt{Ent}$ and models that predict $\log_{10}(Ent)$ were evaluated. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed

values) as compared to computed Ent were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values). This comparison lead to the conclusion that the most appropriate and reliable model would be one that estimated $\log_{10}$(Ent).

Turb and seasonality were selected as the best predictors of Ent based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2) and relatively low model standard percentage error (*MSPE*), prediction error sum of squares (PRESS), and Mallow's C_p. Values for all of the afore mentioned statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

Model Summary

Summary of final regression analysis for Ent concentration at site number 06892350.

Ent concentration-based model:

$$\log_{10}(Ent) = 1.39 \times \log_{10}(Turb) + 0.211 \times \sin\left(\frac{2\pi DY}{365}\right) + 0.214 \times \cos\left(\frac{2\pi DY}{365}\right) - 0.292$$

where

Ent = Enterococci in in colonies per 100 milliliter (col/mL);

Turb = turbidity in formazin nephelometric units (FNU);

Sin & Cos = seasonality component; and,

DY= day of the year.

Turbidity and seasonality make physical and statistical sense as explanatory variables for Ent.

The log-transformed model may be retransformed to the original units so that Ent can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's Bias Correction Factor (BCF). For this model, the calculated BCF is 2.23. The retransformed model, accounting for BCF is:

$$Ent = Turb^{1.39} \times 10^{0.211 \sin\left(\frac{2\pi DY}{365}\right)} \times 10^{0.214 \cos\left(\frac{2\pi DY}{365}\right)} \times 1.14$$

Previous Models

Start year	End year	Model
2000	2005	$\log_{10}(Ent) = 1.64 \times \log_{10}(Turb) - 0.768$

Enterococci Concentration Record

The Ent record is computed using this regression model and stored at the National Real-Time Water Quality (NRTWQ) Web site. Data are computed at 15-minute intervals. The complete water-quality record can be found at http://nrtwq.usgs.gov/ks.

Remarks

None

R Output for Enterococci; 06892350; Kansas River at De Soto, KS

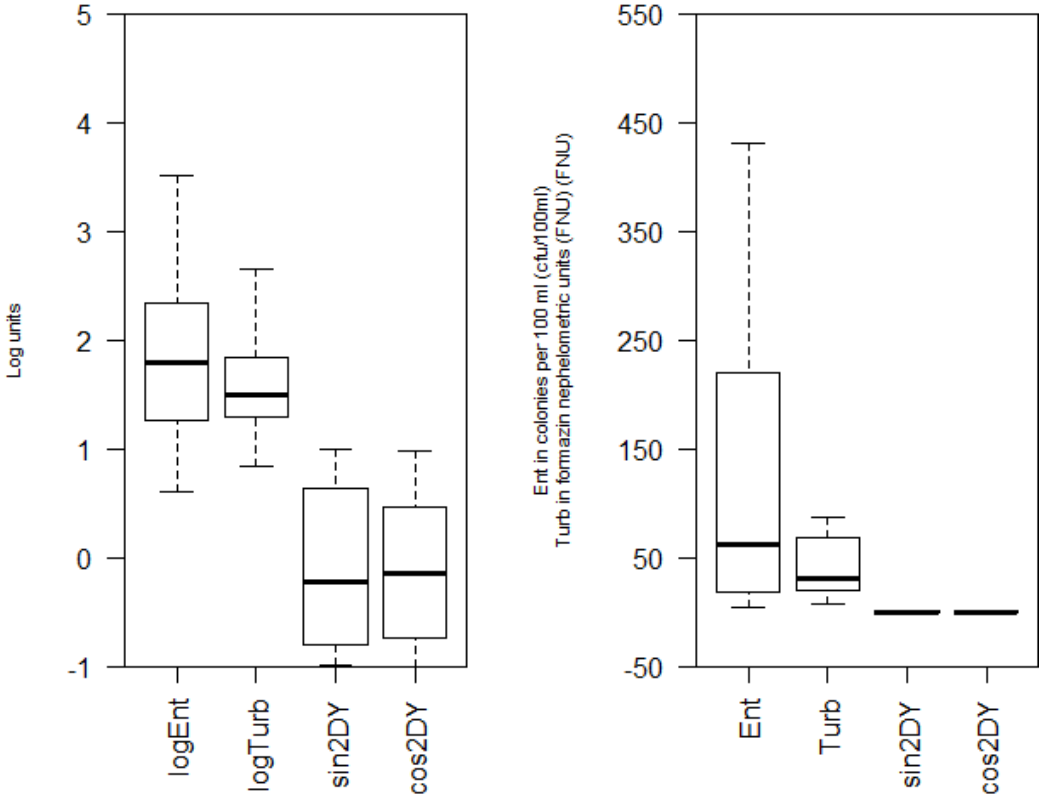
Model Form

logEnt = + 1.39 * logTurb + 0.211 * sin2DY + 0.214 * cos2DY + -0.292

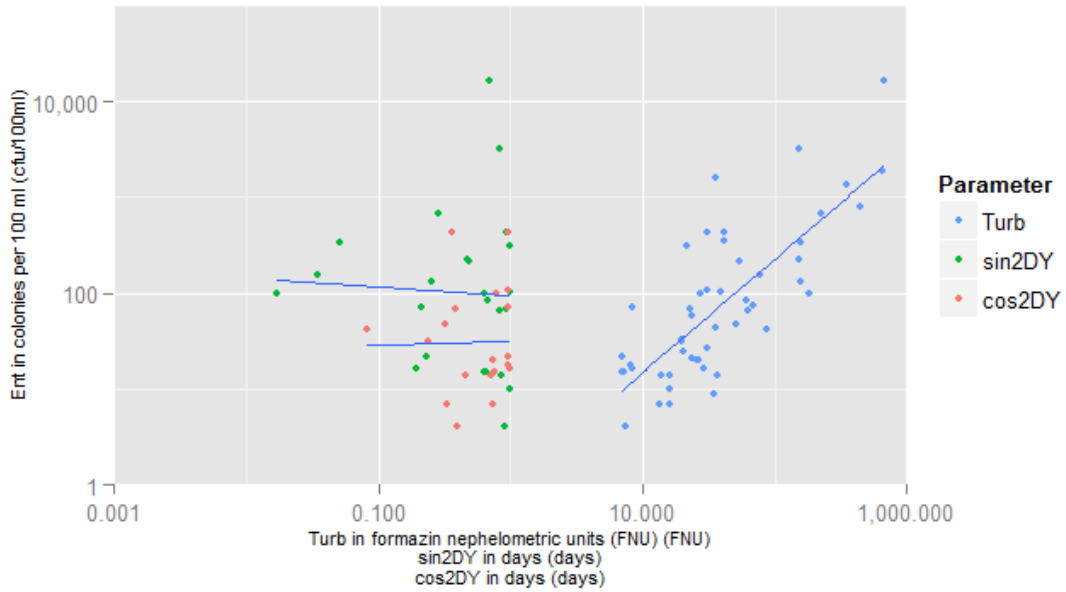
Variable Summary Statistics

	logEnt	logTurb	sin2DY	cos2DY	Ent	Turb
Minimum	0.602	0.845	-0.9970	-1.000	4	7.0
1st Quartile	1.260	1.290	-0.8020	-0.746	18	19.7
Median	1.790	1.490	-0.2260	-0.152	62	31.0
Mean	1.870	1.590	-0.0965	-0.112	601	87.8
3rd Quartile	2.340	1.840	0.6400	0.460	220	69.0
Maximum	4.220	2.830	0.9970	0.981	16700	674.0

Box Plot(s) of sample data



Exploratory Plot



Model Calibration

Basic Data

Number of Observations	50
Standard error (RMSE)	0.469
Upper Model standard percentage error (MSPE)	194
Lower Model standard percentage error (MSPE)	66
Coefficient of determination (R^2)	0.661
Adjusted Coefficient of Determination (Adj. R^2)	0.639
Bias Correction Factor (BCF)	2.23

```
Variance Inflation Factors (VIF)
 logTurb   sin2DY   cos2DY
1.805221 1.002415 1.802108
```

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	-0.292	0.2770	-1.05	2.99e-01
logTurb	1.390	0.1750	7.92	3.94e-10
sin2DY	0.211	0.0935	2.26	2.86e-02
cos2DY	0.214	0.1290	1.66	1.04e-01

Correlation Matrix

	Intercept	logTurb	sin2DY	cos2DY
Intercept	1.0000	-0.9700	-0.0144	-0.6180
logTurb	-0.9700	1.0000	0.0479	0.6670
sin2DY	-0.0144	0.0479	1.0000	0.0238
cos2DY	-0.6180	0.6670	0.0238	1.0000

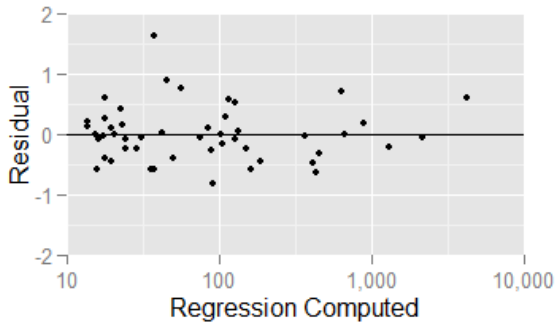
Test Criteria

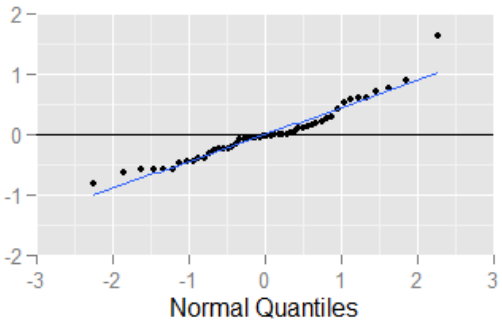
Leverage	Cook's D	DFFITS
0.1800000	0.2633403	0.4898979

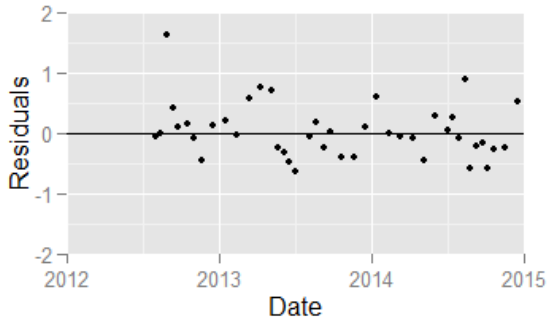
Flagged Observations

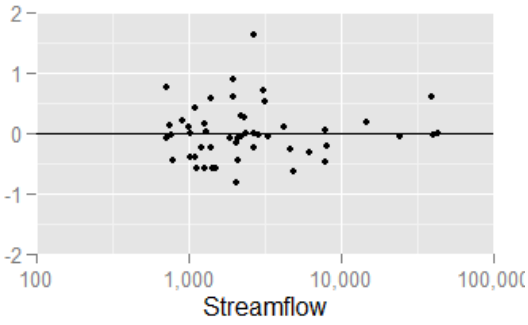
	logEnt	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
8/27/2012 8:25	3.196	1.571	1.6240	3.573	4.158	0.06032	0.20490	1.0540
8/11/2014 15:10	2.544	1.657	0.8868	1.953	2.018	0.06300	0.06414	0.5232
5/4/2015 13:30	1.146	1.958	-0.8119	-1.794	-1.840	0.06866	0.05932	-0.4996
5/18/2015 15:30	4.222	3.626	0.5954	1.410	1.426	0.18930	0.11610	0.6889

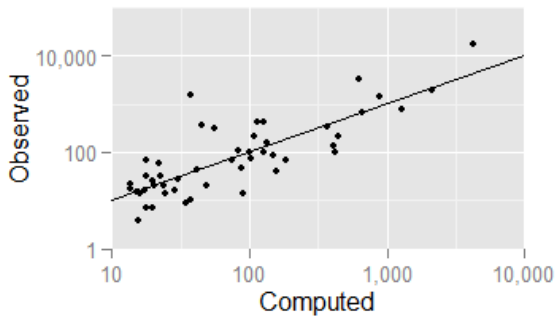
Statistical Plots

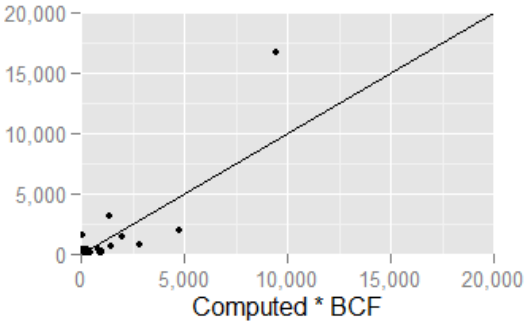


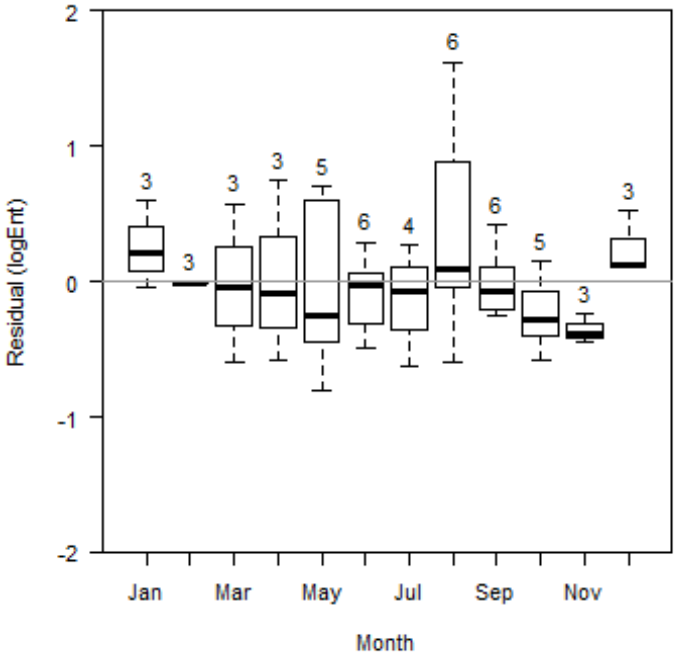


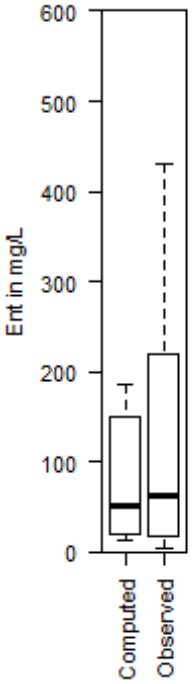


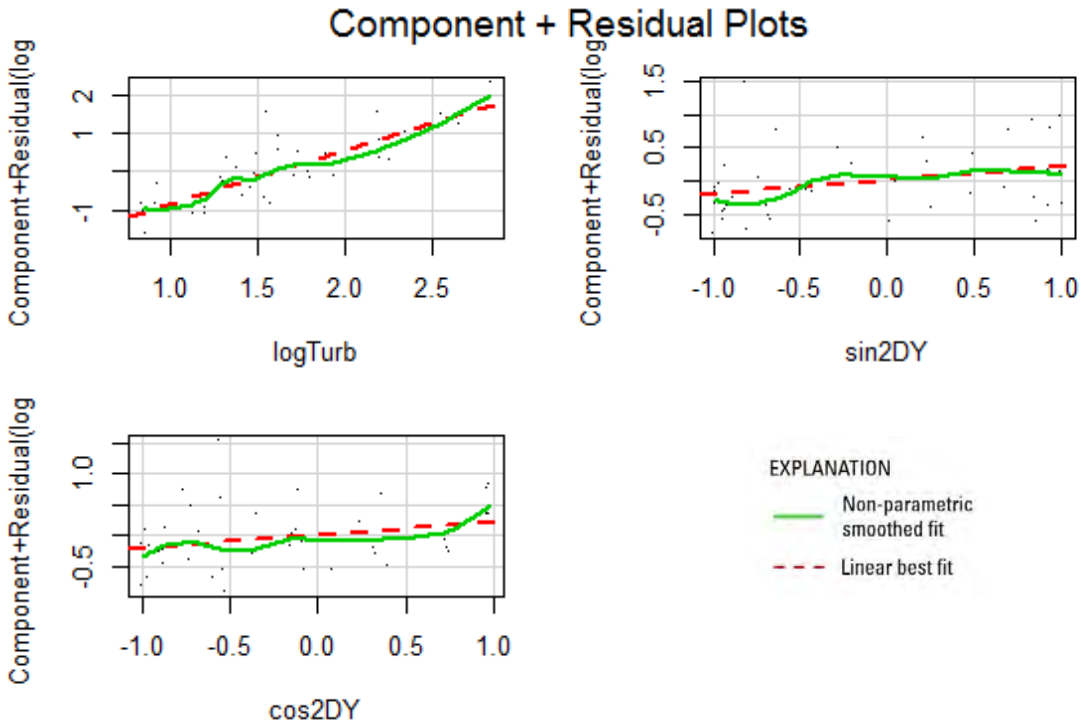







Residual (logEnt)
2
1
0
-1
-2
3
3
3
3
5
6
4
6
6
5
3
3
Jan
Mar
May
Jul
Sep
Nov
Month


Ent in mg/L
600
500
400
300
200
100
0
Computed
Observed


Component + Residual Plots
Component+Residual(log
2
1
-1
1.0
1.5
2.0
2.5
logTurb
Component+Residual(log
1.5
0.5
-0.5
-1.0
-0.5
0.0
0.5
1.0
sin2DY
Component+Residual(log
1.0
-0.5
-1.0
-0.5
0.0
0.5
1.0
cos2DY
EXPLANATION
Non-parametric smoothed fit
Linear best fit

Models considered

Model Formula	Number of Variables	Standard Error	R2	Adjusted R2	Cp	PRESS	VIF	MSPE
logEnt ~ logTurb	1	0.4959	60.43	59.61	5.12	12.78	1	± 140
logEnt ~ Turb	1	0.5529	50.82	49.8	17.54	16.14	1	± 160
logEnt ~ logQ	1	0.5657	48.53	47.45	20.5	16.89	1	± 170
logEnt ~ logTurb + sin2DY	2	0.4776	64.07	62.54	2.419	12.13	1.002	± 130
logEnt ~ logTurb + cos2DY	2	0.489	62.33	60.72	4.672	12.58	1.801	± 140
logEnt ~ logQ + logTurb	2	0.4922	61.83	60.21	5.309	12.78	2.853	± 140
logEnt ~ logTurb + sin2DY + cos2DY	3	0.469	66.09	63.88	1.806	11.85	1.805	± 130
logEnt ~ Turb + logTurb + sin2DY	3	0.4748	65.24	62.98	2.904	12.22	3.29	± 130
logEnt ~ Q + logTurb + sin2DY	3	0.4792	64.6	62.29	3.737	12.46	1.799	± 130
logEnt ~ Q + logTurb + sin2DY + cos2DY	4	0.4713	66.5	63.52	3.288	12.09	1.807	± 130
logEnt ~ Turb + logTurb + sin2DY + cos2DY	4	0.4714	66.49	63.51	3.3	12.26	3.746	± 130
logEnt ~ logQ + logTurb + sin2DY + cos2DY	4	0.4716	66.46	63.48	3.331	12.18	3.075	± 130
logEnt ~ logQ + Turb + logTurb + sin2DY + cos2DY	5	0.4752	66.7	62.92	5.018	12.59	3.272	± 130
logEnt ~ Q + Turb + logTurb + sin2DY + cos2DY	5	0.4754	66.67	62.88	5.063	12.37	2.109	± 130
logEnt ~ Q + logQ + logTurb + sin2DY + cos2DY	5	0.4764	66.53	62.73	5.241	12.66	4.334	± 130
logEnt ~ Q + logQ + Turb + logTurb + sin2DY + cos2DY	6	0.4806	66.72	62.07	7	12.97	4.779	± 130

Data

	Date	logEnt	logTurb	sin2DY	cos2DY	Ent	Turb	Computed logEnt	Computed Ent	Residual	Normal Quantiles	Censored Values
1	2012-07-30	1.431	1.488	-0.4705	-0.8824	27	30.79	1.484	68.15	-0.053	-0.176	--
2	2012-08-13	1.312	1.371	-0.6656	-0.7463	20.5	23.5	1.31	45.56	0.00217	0.227	--
3	2012-08-27	3.196	1.556	-0.823	-0.568	1570	36	1.571	83.27	1.62	2.26	--
4	2012-09-10	1.763	1.38	-0.9355	-0.3533	58	24	1.349	49.94	0.414	0.949	--
5	2012-09-24	1.398	1.314	-0.9928	-0.1195	25	20.59	1.295	44.07	0.103	0.438	--
6	2012-10-15	1.505	1.301	-0.9712	0.2384	32	20	1.359	51.04	0.146	0.61	--
7	2012-10-29	1.134	1.146	-0.888	0.4598	13.6	14	1.209	36.15	-0.0755	-0.227	--
8	2012-11-19	0.8451	1.13	-0.6668	0.7452	7	13.5	1.295	44.06	-0.45	-0.949	--
9	2012-12-17	1.255	0.9165	-0.2467	0.9691	18	8.25	1.135	30.49	0.12	0.551	--
10	2013-01-14	1.342	0.8451	0.2296	0.9733	22	7	1.138	30.67	0.205	0.735	--
11	2013-02-11	1.176	0.8579	0.6544	0.7562	15	7.21	1.199	35.3	-0.0227	0.075	--
12	2013-03-11	2.633	1.495	0.9298	0.3681	430	31.28	2.058	255.2	0.576	1.12	--
13	2013-04-08	2.491	1.332	0.9944	-0.106	310	21.5	1.744	123.9	0.747	1.63	--
14	2013-05-06	3.505	2.188	0.8314	-0.5557	3200	154	2.799	1407	0.706	1.46	--
15	2013-05-20	1.919	1.788	0.675	-0.7378	83	61.4	2.173	332.9	-0.254	-0.61	--
16	2013-06-03	2.342	2.184	0.4786	-0.878	220	152.9	2.651	1000	-0.309	-0.735	--
17	2013-06-17	2.124	2.204	0.2557	-0.9668	133	160	2.613	915.3	-0.489	-1.12	--
18	2013-07-01	2	2.258	0.01697	-0.9999	100	181	2.629	951.2	-0.629	-1.85	--
19	2013-08-05	3.279	2.822	-0.5515	-0.8341	1900	663.3	3.327	4743	-0.0483	-0.125	--
20	2013-08-19	3.136	2.551	-0.734	-0.6792	1367	356	2.947	1976	0.189	0.671	--
21	2013-09-09	1.204	1.459	-0.9269	-0.3754	16	28.75	1.455	63.73	-0.251	-0.551	--
22	2013-09-23	1.643	1.554	-0.9896	-0.1439	44	35.84	1.624	94.08	0.019	0.33	--
23	2013-10-21	0.8451	1.204	-0.9434	0.3318	7	16	1.25	39.76	-0.405	-0.873	--
24	2013-11-18	1.301	1.422	-0.6797	0.7335	20	26.4	1.694	110.4	-0.393	-0.802	--
25	2013-12-16	2.029	1.489	-0.2664	0.9639	107	30.8	1.923	187.3	0.106	0.494	--
26	2014-01-13	1.845	0.9294	0.2107	0.9775	70	8.5	1.252	39.87	0.593	1.22	--
27	2014-02-10	1.176	0.8482	0.6397	0.7687	15	7.05	1.185	34.19	-0.00873	0.125	--
28	2014-03-10	1.826	1.362	0.9245	0.3812	67	23	1.874	167.2	-0.0482	-0.075	--
29	2014-04-07	2.013	1.591	0.996	-0.0893	103	39	2.107	285.6	-0.0938	-0.33	--
30	2014-05-05	1.82	1.803	0.8402	-0.5422	66	63.57	2.271	416.9	-0.451	-1.03	--
31	2014-06-02	2.328	1.739	0.4938	-0.8696	213	54.83	2.039	244.1	0.29	0.873	--
32	2014-06-30	2.185	1.892	0.0349	-0.9994	153	78	2.126	298.6	0.0586	0.384	--
33	2014-07-14	1.519	1.294	-0.205	-0.9788	33	19.67	1.25	39.72	0.269	0.802	--
34	2014-07-28	1.301	1.411	-0.4333	-0.9012	20	25.75	1.381	53.65	-0.0795	-0.278	--
35	2014-08-11	2.544	1.621	-0.6366	-0.7712	350	41.8	1.657	101.5	0.887	1.85	--
36	2014-08-25	0.9542	1.539	-0.8019	-0.5974	9	34.6	1.546	78.48	-0.591	-1.63	--
37	2014-09-08	2.903	2.652	-0.9209	-0.3899	800	449	3.109	2871	-0.206	-0.438	--

38	2014-09-22	1.863	1.839	-0.987	-0.1607	73	69	2.016	231.8	-0.153	-0.384	--
39	2014-10-06	1.613	1.936	-0.9966	0.08203	41	86.33	2.201	354.8	-0.588	-1.33	--
40	2014-10-20	1.672	1.711	-0.9481	0.3179	47	51.4	1.949	198.8	-0.277	-0.671	--
41	2014-11-17	1.146	1.204	-0.6933	0.7207	14	16	1.387	54.4	-0.24	-0.494	--
42	2014-12-15	2.633	1.623	-0.2816	0.9595	430	41.99	2.106	285.1	0.528	1.03	--
43	2015-01-12	1.204	0.9227	0.1963	0.9805	16	8.37	1.24	38.81	-0.0358	0.025	--
44	2015-02-09	2	1.439	0.6298	0.7768	100	27.5	2.004	225.7	-0.00442	0.176	--
45	2015-03-09	0.6021	0.8692	0.9174	0.398	4	7.4	1.193	34.86	-0.591	-1.46	--
46	2015-04-06	1	1.206	0.9972	-0.07441	10	16.07	1.576	84.15	-0.576	-1.22	--
47	2015-05-04	1.146	1.574	0.8496	-0.5274	14	37.49	1.958	202.8	-0.812	-2.26	--
48	2015-05-18	4.222	2.829	0.6982	-0.7159	16670	674.5	3.626	9452	0.595	1.33	--
49	2015-06-15	2.824	2.348	0.2872	-0.9579	667	223	2.821	1479	0.00316	0.278	--
50	2015-06-29	2.519	2.204	0.05198	-0.9986	330	160	2.563	815.9	-0.0441	-0.025	--

Definitions and National Water Information System (parameter code)

Ent: Enterococci in cfu/100mL (90909)
Turb: Turbidity in FNU (63680)
DY: Date in decimal years